User-centric Evolutionary Design Systems - the Visualisation of Emerging Multi-Objective Design Information

J.A.R Abraham, I.C. Parmee
Advanced Computation in Design and Decision-making (ACDDM)
http://www.ad-comtech.co.uk/ACDDM_Group.htm
University of the West of England, Bristol
johnson.abraham@uwe.ac.uk ian.parmee@uwe.ac.uk

Summary

The paper describes further developments of the interactive evolutionary design concept relating to the emergence of mutually inclusive regions of high performance design solutions. These solutions are generated from cluster-oriented genetic algorithm (COGAs) output and relate to a number of objectives introduced during the preliminary design of military airframes. The data-mining of multi-objective COGA (moCOGA) output further defines these regions through the application of clustering algorithms, data reduction and variable attribute relevance analyses. A number of visual representations of the COGA output projected onto both variable and objective space are presented. The multi-objective output of the COGA is compared to output from a Strength Pareto Evolutionary Algorithm (SPEA-II) to illustrate the manner in which moCOGAs can generate good approximations to Pareto frontiers.

1 Introduction

In recent years research relating to user-centric evolutionary search and exploration approaches in engineering domains has become established and complete, partial and more implicit human judgement has been introduced to assist machine-based design evaluation and representation. Although entirely machine-centred optimisation has significant utility within well-defined routine and detailed design domains more designer-centred evolutionary search and exploration can play a major role during early design where initial poor definition and uncertainty are inherent features. The paper concentrates upon the interactive evolutionary design (IED) concept, a human centric approach (Parmee, 2002; Parmee et al, 2000) supporting the generation and extraction of high-quality design information during the early stages of design. The aim is the provision and succinct presentation of information appertaining to complex relationships between the variables, objectives and constraints that define a developing design space.

Recent work complements the IED concept by further attempting to meld experiential knowledge and intuition with powerful machine-based search, exploration and information processing. Graphical representations of variable and objective space provide a variety of perspectives illustrating the relationships between them (Parmee and Abraham, 2003). This information emerges from cluster-oriented genetic algorithm (COGA) output and is further defined by appropriate data mining, processing and visualization techniques (Abraham and Parmee 2004a). The intention is to support implicit learning and reduce complexity by supporting the designer in the development of both a quantitative and intuitional understanding of the problem domain (Parmee and Abraham 2004b)

2 COGAs and The BAE Systems MiniCAPs Model

Cluster Oriented Genetic Algorithms were developed in the early 1990s to provide a means to identify high-performance (HP) regions of complex conceptual design spaces and enable the extraction of information from such regions relating, initially, to solution sensitivity (Parmee 1996). COGAs identify HP regions through the on-line adaptive filtering of solutions generated by a genetic algorithm. Further work resulted in several variations of COGA and also identified and illustrated the manner in which the COGA approach can be utilised to generate highly relevant design information relating to single, multi-objective and constrained problem domains (Bonham & Parmee, 2004; Parmee and Bonham, 1999).

1. Climb Mach Number (CLMN)	4. Gross Wing Plan Area (GWP)	7. Wing Lead Edge Sweep (WLES)
2. Cruise Height (CH)	5. Wing Aspect Ratio (WAR)	8. Wing T/C Ratio (WTCR)
3. Cruise Mach Number (CRMN)	6. Wing Taper Ratio (WTR)	9. By Pass Ratio (BPR)

Table 1
MiniCAPS Input Variables

COGA comprises two primary components: the diverse search engine which utilises a genetic algorithm to search the design space identifying regions of high performance relating to a particular objective and the adaptive filter (AF) which extracts and stores information relating to each identified region. The Adaptive Filter (AF) copies high fitness solutions from the evolving population to the Final Clustering Set (FCS). The user can vary the severity of the filtering mechanism in order to identify regions ranging from succinct groupings of very high performance solutions to larger regions of high and lower performance solutions. Sufficient regional set-cover (in terms of number of solutions) can be achieved to allow significant qualitative and quantitative design information to be extracted. COGA development and application has been well documented and is widely referenced within the text. Many of the COGA and IEDS papers referenced can be downloaded from: http://www.ad-comtech.co.uk/Parmee-Publications.htm.

Earlier IED research has utilised the BAE Systems MiniCAPs model, a simplified version of the CAPS (Computer Aided Project Studies) suite of preliminary design models for the early investigation stages of military aircraft airframe design. MiniCAPS (Webb 1997) was initially developed for research purposes relating to the development of the IED concept. The MiniCAPS model comprises nine continuous input variables and twelve continuous output parameters. Subroutines calculate properties relating to criteria such as performance, wing geometry, propulsion, fuel capacity, structural integrity etc. Input variables are listed in Table 1. The recent research described in the paper has again initially utilized MiniCAPS although the techniques described and variations of them are also currently being applied in the conceptual design of submersible vehicles and in pharmaceutical drug design and discovery.

3 Identifying High-performance Regions Relating to Differing Objectives

Figures 1a, b & c show HP regions comprising solutions from the FCSs relating to three of the twelve MiniCAPS objectives (Ferry Range (FR), Attained Turn Rate (ATR1) and Specific Excess Power (SEP1)) projected onto a variable hyperplane relating to two of the nine variables utilized in the search process. This projection allows the designer to visualize the HP regions, identify their bounds and subsequently reduce the variable ranges as described in previously referenced papers. These papers also introduce the projection of these differing objective HP regions onto the same variable hyperplane as shown in figure 2 from which the degree of objective conflict immediately becomes apparent to the designer. The emergence of a mutually inclusive region of HP solutions relating to the ATR1 and FR objectives indicates a low degree of conflict whereas the HP region relating to SEP1 is remote (in variable space) to both the ATR1 and FR regions indicating a higher degree of conflict. The Adaptive Filter setting has been kept constant across the COGA runs relating to each objective.

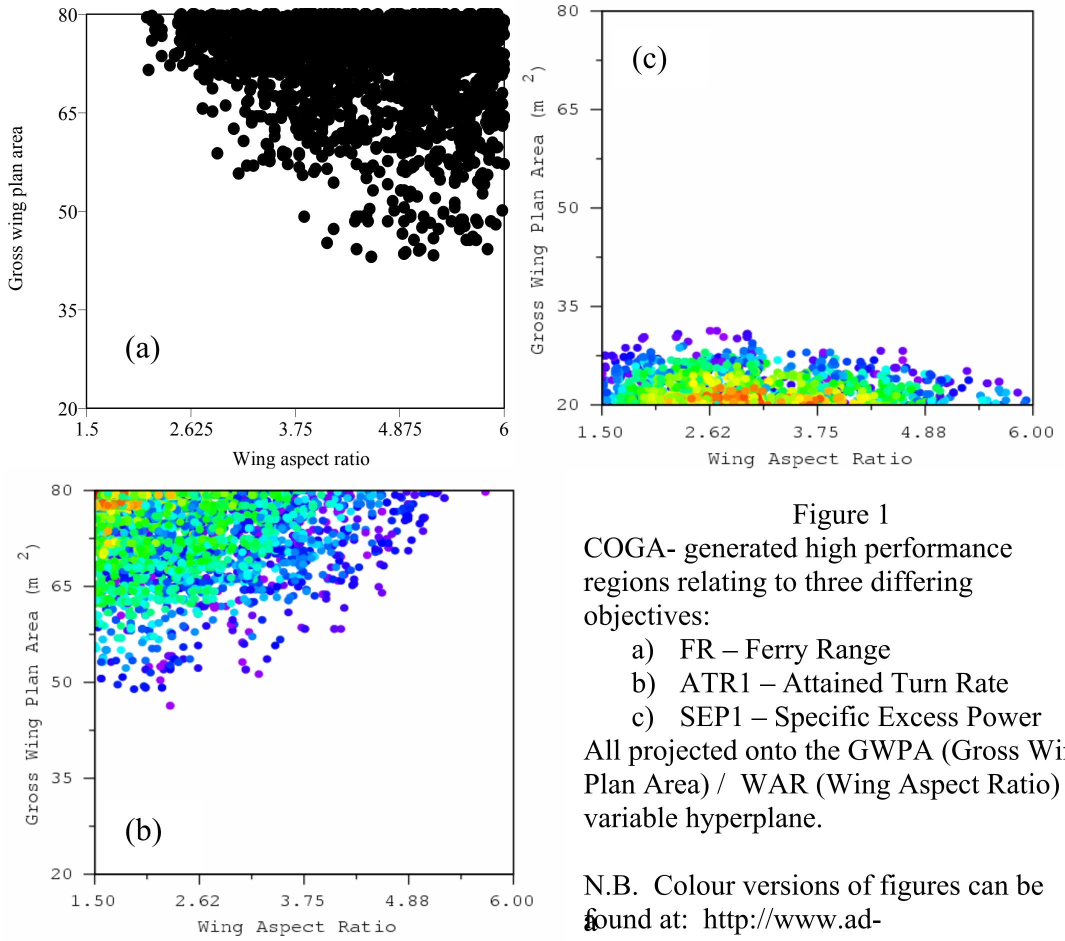

Figure 1
COGA- generated high performance regions relating to three differing objectives:

a) FR – Ferry Range
b) ATR1 – Attained Turn Rate
c) SEP1 – Specific Excess Power

All projected onto the GWPA (Gross Wing Plan Area) / WAR (Wing Aspect Ratio) variable hyperplane.

N.B. Colour versions of figures can be found at: http://www.ad-comtech.co.uk/cogaplots.htm

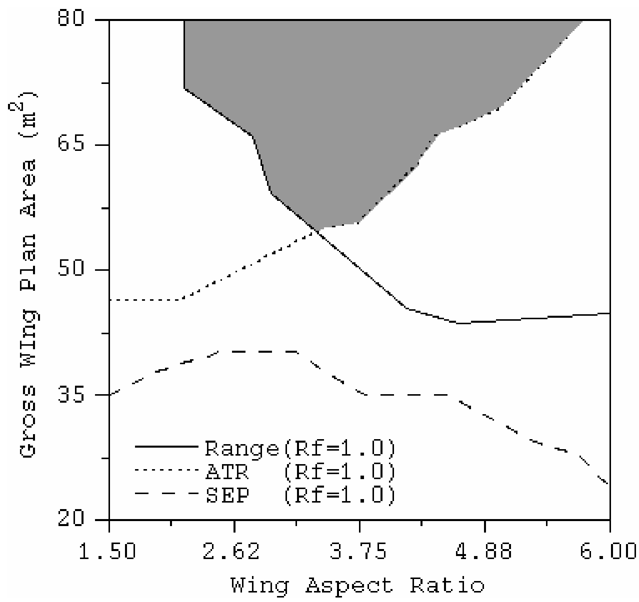

Figure 2
All HP regions projected on to the GWPA / WAR variable hyperplane.

There is a good deal of information contained in the FCS solution sets / HP regions. This relates to appropriate variable ranges for single objectives, degree of conflict between multiple objectives and the emergence and definition of mutually inclusive (common) regions. Although such graphical representation provides an excellent spatial indication of the degree of objective conflict searching through all two dimensional variable hyperplanes to visualise such information is not a feasible approach. Recent research has resulted in single graphical representations that can present all variable and objective data whilst providing easily utilised links to other visual perspectives. The Parallel Co-ordinate box plot representation shown in figure 3 is one such graphic that provides a central repository containing much relevant single and multiple-objective solution information.

4 Parallel Co-ordinate Box Plots (PCBP)

Parallel Co-ordinate plots (Inselberg 1985) appeared to offer best potential in terms of providing a single graphic illustrating complex relationships between variable and objective space. Parallel Co-ordinate representation displays each variable dimension vertically parallel to each other. Points corresponding to a solution's value of that variable can then be plotted on each vertical variable axis. It is thus possible to show the distribution of solutions in all variable dimensions and the correlation between different dimensions. The disadvantage of the technique when attempting to include multiple objectives is that the density of the information presented hinders perception. To overcome the 'data density' problem three modifications to the standard Parallel Co-ordinate representation have been included:

i) additional vertical axes for each variable so that each objective can be represented;
ii) an indication of the degree of HP region solution cover across each variable range;
iii) the introduction of Box Plots to indicate skewness of solutions across each variable range.

This Parallel Co-ordinate Box Plot (PCBP) representation provides a much clearer graphic as shown in figure 3. The vertical axis of each variable plane is scaled between the minimum and maximum value of the variable found in the HP region / FCS of each particular objective i.e. the length of the axis represents the normalized ranges of variable values present in a HP region. If an objective HP solution set does not extend across the whole of the variable range the axis is terminated by a whisker at the maximum or minimum value of the variable. The colour-coded box plots relate to each objective (i.e. SEP1, ATR1 and FR). The median is marked within the box and the box extends between the lower and upper quartile values within the variable set. The PCBP clearly visualizes the skewness of solution distribution relating to each objective in each variable dimension. Differing degrees of skewness provide an indication of the degree of conflict between objectives.

For instance, it is immediately apparent that all three objective boxes largely overlap in the case of variables 1, 2, 3, 6 and 9. However, significant spatial differences in the distribution of the boxes are evident in terms of at least one objective where variables 4, 5, 7, and 8 are concerned. Referring back to Table 1, variables 4 and 5 are Gross Wing Plan Area and Wing Aspect Ratio. The conflict between SEP1 and FR / ATR1 evident in figure 2 is strongly reflected in the HP solution distribution indicated by the whisker truncation of variable 4 in figure 3 and in the box plots of that variable. In terms of variable 5 the whisker terminations relating to ATR1 and FR in figure 3 reflect the extent of the solution distribution across their HP regions in figure 2. The box plots also reflect the relative distribution of HP solutions of all objectives along that variable plane as illustrated in figure 2.

Figure 4 shows a projection of the ATR1 HP region onto the Cruise Hieght (variable 1) and Climb Mach No (variable 2) hyperplane The relatively uniform distribution of HP solutions across the hyperplane is reflected in the appropriate variable plots of figure 3.

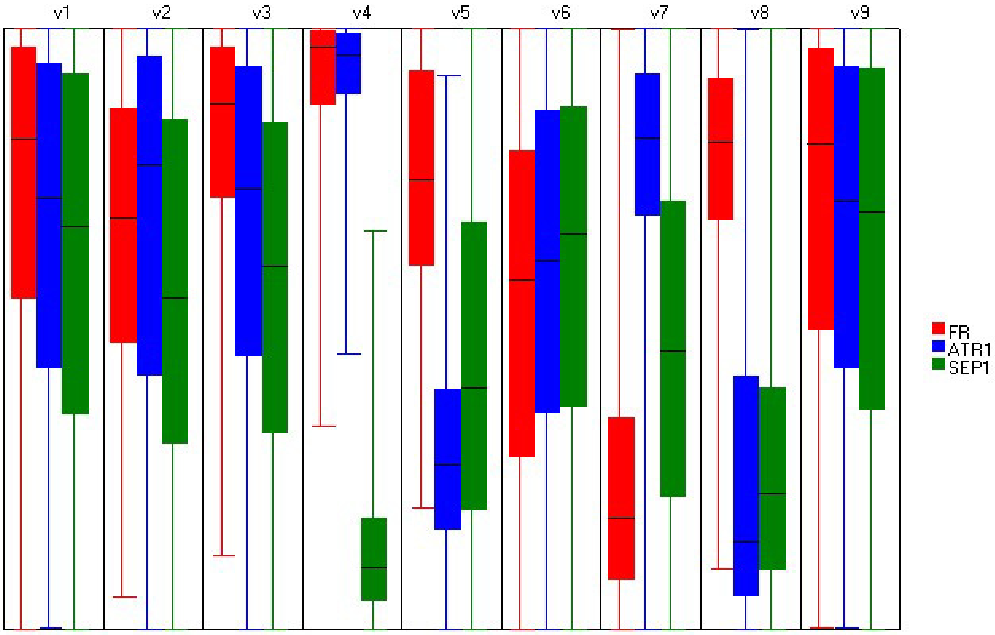

Figure 3
Parallel Box Plot of solution distribution of each objective across each variable dimension

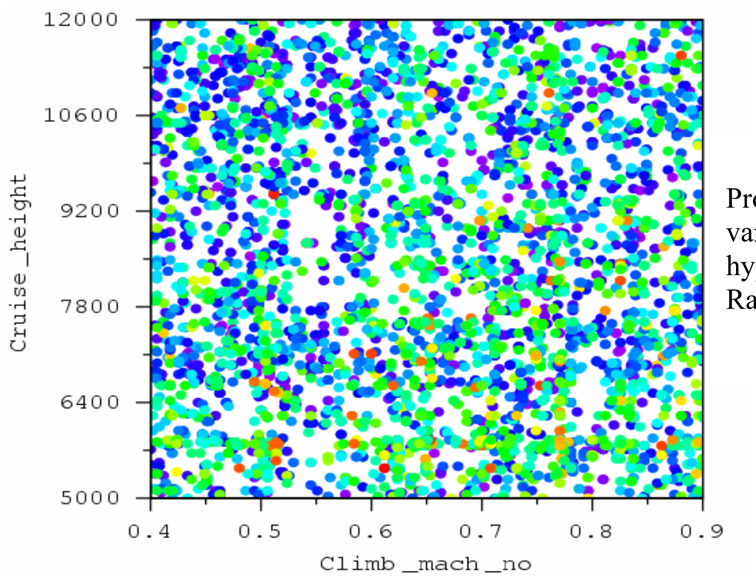

Figure 4
Projection of results onto variable 1 / variable 2 hyperplane for Attained Turn Rate (ATR1) objective

Extensive variable attribute relevance analysis (Kamber 2001; Arvin and Langley 1997) utilising the COGA-generated HP solutions has been carried out in addition to standard skewness calculations to verify the visual information available in the PCBP (Abraham and Parmee 2004). Variable attribute relevance analysis quantifies the relevance of an attribute (i.e. variable) with respect to a given class or concept by measures such as information gain and correlation co-efficient. Using these procedures the information gain of each variable is calculated and variables are ranked in terms of the degree of effect they have across the set of three objectives. Results are shown in Table 2. The ranking identifies variables 4,5,7 and 8 as those variables to which the objective set is most sensitive. This, plus skewness analysis, confirms the visual information available in the plot.

Input Variable	Skewness			Correlation Coefficient			Inform-ation Gain	Rank
	ATR1	FR	SEP1	ATR1	FR	SEP1	ATR1, FR& SEP1	
1. CLMN	-0.481	-0.888	0.013	0.095	0.136	-0.086	0.026	7
2. CH	-0.566	-0.193	-0.430	0.059	0.307	0.043	0.068	6
3. CMN	-0.475	-1.123	-0.151	0.051	0.181	0.049	0.118	5
4. GWPA	-1.653	-1.758	1.280	0.170	0.463	-0.566	0.953	1
5. WAR	0.501	-0.404	0.761	-0.257	0.251	-0.207	0.255	4
6. WTR	-0.230	0.172	-0.008	0.013	0.001	-0.018	0.013	9
7. WLES	-1.351	1.098	0.315	0.478	-0.349	-0.071	0.265	3
8. WTCR	1.059	-0.922	1.073	-0.55	0.249	-0.521	0.419	2
9. BPR	-0.460	-0.757	-0.127	0.141	0.119	0.019	0.014	8
Mean of Information Gain							0.237	

Table 2
Results of Skewness and Attribute Relevance Analysis

5 Exploring the Relationship Between COGA and MOGA Output.

If we take the FCS / HP region solutions for ATR1 and FR and plot them in objective space the distributions shown in figure 5 emerge. We have always assumed a relationship between the solutions in the FCSs and a Pareto frontier and the outer edge of the plot would seem to support this assumption (Parmee and Abraham 2003; Abraham and Parmee 2004). The working principle of multi-objective moCOGAs is very different to that of standard evolutionary multi-objective algorithms (Deb 2001) which tend to use a non-dominance Pareto-based approach.

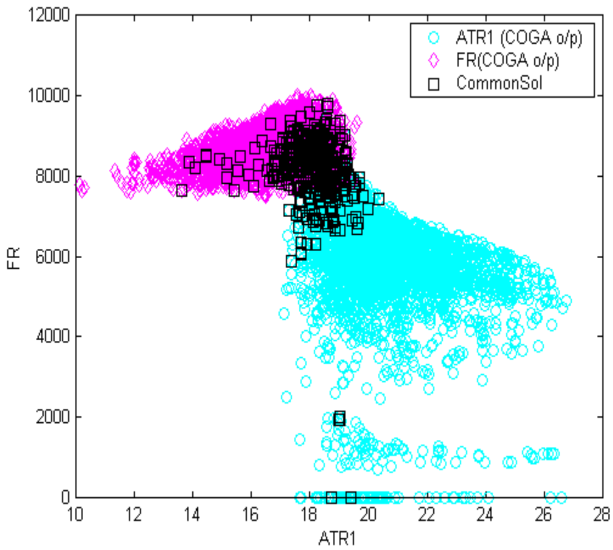

Figure 5
Distribution of FR and ATR1 solutions in objective space

The principle of moCOGA is to generate as much information as possible concerning high performance regions relating to various objectives within a problem space. Using a standard multi-objective GA (MOGAs) it is possible to obtain solutions lying along the Pareto front but difficult to explore the relationship between variable and the objective space and to discover what is occurring close to the frontier. During the early stages of design it is quite possible that

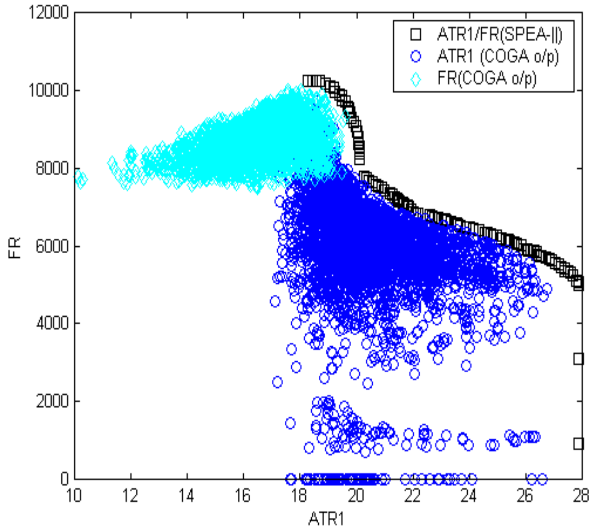

Figure 6
Distribution of ATR1 and FR solutions against SPEA-II Pareto front.

the designer is also interested in such solutions and solutions that lie around particular sections of the Pareto front.

Most multi-objective genetic algorithms use the concept of Pareto dominance. Deb and Ziztler (2001) compared a set of these algorithms and results have shown that the Strength ParetoEvolutionary Algorithm (SPEA – Zitzler et al 2002) performs comparatively well. SPEA's strength lies in its use of elitism (the concept of storing and using the good solutions in earlier generations for future search). The SPEA-II algorithm has been utilised to generate Pareto fronts for the objectives SEP1, ATR1 and FR for comparative purposes. Figure 6 illustrates the distribution of COGA output and the SPEA-II output in objective space. Similar figures relating to all three objectives can be found in Abraham and Parmee (2004); Parmee and Abraham (2004). The moCOGA approach therefore provides a good visual indication of the degree of conflict between objectives; an opportunity to explore varying objective preferences and view their effect upon HP region bounds and the abilty to generate an approximate Pareto front relating to the objectives under investigation plus solutions around the Pareto front. This is in addition to the utility of COGA in single objective space as described in previous papers.

6 Common Region Identification

Each COGA run generates a final clustering set (FCS) containing HP solutions for that particular objective. The solutions in each FCS form a cluster with particular characteristics defining the objective. Identifying the common region between all objectives' FCSs is a classification problem where classified sets of solutions share the neighbourhood with solutions satisfying other objectives. There is some utility in identifying the bounds of this common region and extracting the HP solutions from it for further analysis.

The K-Nearest Neighbour, KNN, classifier (Han and Kamber 2001) is adopted but, in this instance, the technique is used to classify solutions from all FCSs into a cluster where each solution relating to a particular objective shares a common neighborhood with solutions belonging to other objectives' FCSs. Initially the FCS solutions of two separate COGA runs relating to objectives ATR1 and FR provided input to the classifier which then identifies those

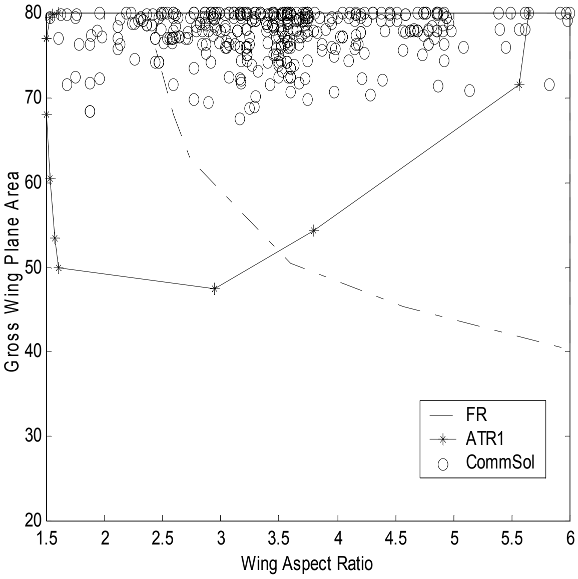

Figure 7
ATR1/FR common region identified using all the solutions and all the variables for K=3

solutions that have near neighbours belonging to the other objective's FCS. All the nine variables have been utilized in the near-neighbour calculations with K=3. Figure 7 shows that although KNN successfully identifies the common region, outliers are evident. It was assumed that these outliers are caused by the noise introduced into the distance metric by variables that are relatively uncorrelated to the corresponding objective. This assumption is based upon the knowledge that although succinct HP regions exist in some variable hyperplanes, a relatively uniform distribution of the same HP solutions can be evident in others as shown in figure 4.

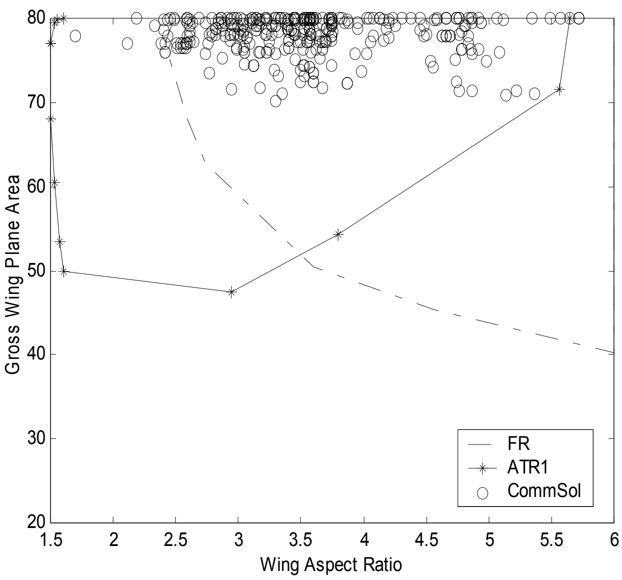

Figure 8
Common Region identified using high information gain variables 5,7 and 8 and all the solutions for K=3

In order to improve clustering efficiency variable attribute relevance analysis results relating to the ATR1 and FR objectives have been utilised to identify highly correlated variables. Input variables with an information gain greater than the mean (5,7 and 8 in the ATR1 / FR case) were then used to identify the common region for K=3. As illustrated in figure 8 this resulted in a reduction in the occurrence of outliers whilst also significantly reducing computational effort. However, as the K-value is increased to improve solution cover in the common region so the number of outliers also increases. This problem can be addressed through the introduction of a data reduction process.

By discarding solutions that are not potential common region solutions computational effort can be further reduced and outliers eliminated even at high K values thereby allowing a higher degree of solution set cover within the identified common region. Data reduction is a rule-based method that removes solutions that are not likely to lie in the common region. The minimum of the rule for each variable is found by calculating the maximum of the minimum values of each variable. The maximum bound of the rule for each variable is found by calculating the minimum of the maximum values of each variable in a particular cluster. Figure 9 shows the standard normal distribution of solutions in the FCSs of objectives FR and ATR1 with respect to WAR, and GWPA. The vertical lines indicate the maximum and minimum bounds of the variables common to all the objectives under consideration (Abraham and Parmee 2004).

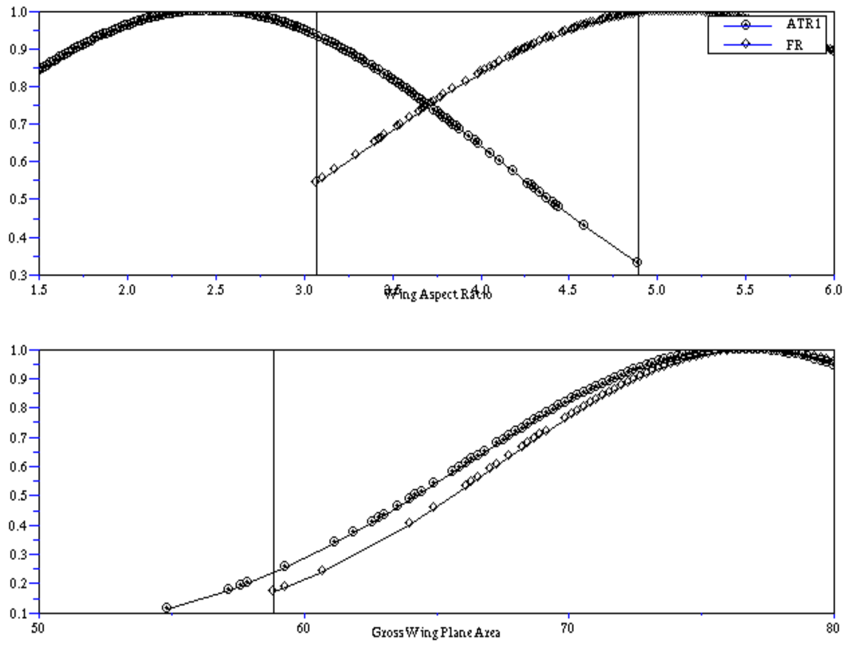

Figure 9
The upper and lower bounds of the variables of Wing Aspect Ratio and Gross Wing Plan Area

Figure 10 illustrates the outcome of the application of the classifier algorithm subsequent to both variable relevance and data reduction. The outlier problem has been eliminated since all the common region solutions identified lie within the line defining the convex hull (Bykat A. 1978) of the objectives ATR1 and FR. Further work is investigating the scaleabilty of these common region identification procedures with respect to increasing numbers of objectives.

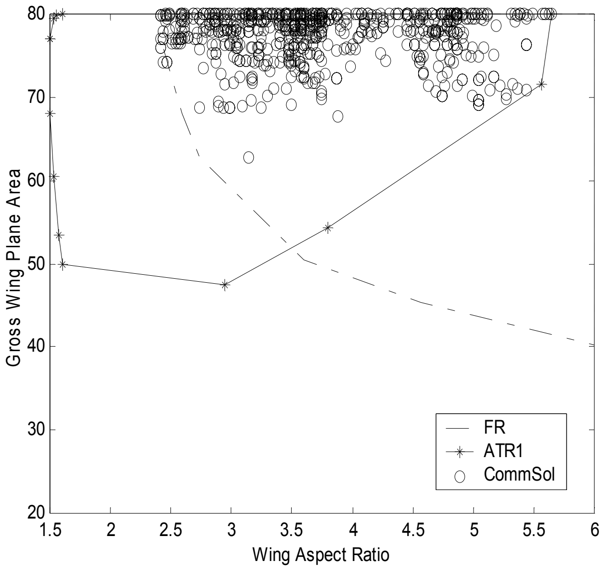

Figure 10
Common Region identified using the relevant, high information gain variables for k=15

7 Conclusion

It has been apparent for some time that moCOGA output can provide a visual representation in variable space of the degree of conflict between objectives as illustrated in figure 2. The designer can interact with COGA to explore how changes to the relative importance of objectives relate to these conflicts. An opportunity for exploration of complex solution relationships across both variable and objective space is therefore available. The introduction of the Parallel Co-ordinate Box Plot representation provides a central repository for this visual information. It is intended that all two dimensional hyperplane graphics will be available to the designer via simple pointing and clicking operations on the relevant axes of the PCBP. Various visual perspectives of variable / objective relationships will therefore be available to the designer to support a clearer understanding of overall multiple objective characteristics.

The ability to generate approximate Pareto frontiers from moCOGA output provides another perspective whilst allowing the designer to explore the potential, in terms of HP solutions, around the non-dominated front. Additional information is therefore available re both qualitative and quantitative objective trade-offs.

Initial results from the mining and processing of data generated from moCOGAs has been presented. Relatively novel usage of established data-mining techniques such as KNN clustering can result in the identification of a mutually inclusive common region in terms of the solution set that describes it. Variable attribute relevance analysis reduces computational overhead by identifying prime variables and eliminating the need for the KNN algorithm to process the full variable set. Data reduction assists this process, significantly lowering computational cost by reducing the number of solutions to be classified. Further work in this area is investigating other methods of on-line identification of common region bounds during moCOGA runs as opposed to waiting for and processing the identified Final Clustering Sets of each objective. Scaleability, in terms of the inclusion of more objectives, is also an area receiving significant attention.

Further detail of the described research can be found in Abraham and Parmee (2004). Cognitive aspects and the development of supporting agent-based systems are discussed in Parmee and Abraham (2004a). The work is further positioned in terms of interactive evolutionary computing and in terms of the initial interactive evolutionary design concept in Parmee and Abraham

(2004b). This book chapter also summarizes the utility of the COGA approach in both single objective and multi-objective design domains.

N.B. Colour versions of figures within the paper can be found at: http://www.ad-comtech.co.uk/cogaplots.htm

References

Abraham J. A. R. and Parmee I. C. (2004). Extraction of Emerging Multi-objective Design Information from COGA Data. Procs. 6th International Conference on Adaptive Computing in Design and Manufacture, Bristol, UK; Springer Verlag (in press).

Bykat A. 1978. Convex Hull of a finite set of points in two dimensions. Informtion Process. Lett.**7**, 296-298.

Arvin L. B. and Langley P. (1997). Selection of Relevant Feature and Examples in Machine Learning", Artificial Intelligence, 10: pp 245-257.

Bonham C. R. and Parmee I. C. (2004) Developments of the Cluster-oriented Genetic Algorithm. Journal of Engineering Optimisation, Taylor and Francis; **36** (2), pp 249 – 279.

Deb K, (2001). Multi Objective Optimization Using Evolutionary Algorithms. John Wiley & Sons.

Inselberg A. (1985). The Plane with Parallel Coordinates. The Visual Computer,**1**, pp 69-91.

Han J. and Kamber M. (2001). Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco, California.

Parmee I.C. and Abraham J. A. R. (2004a). Supporting Implicit Learning via the Visualization of COGA Multi-objective Data. Procs of IEEE Congress on Evolutionary Computation, Portland, Oregon; (in press).

Parmee I.C. and Abraham J. A. R., (2004b), Interactive Evolutionary Design. In: Y. Jin (Ed), Knowledge Incorporation in Evolutionary Computation, Springer Verlag, (in press).

Parmee I.C. and Abraham J. A. R., (2003). Further Developments of the Interactive Evolutionary Design System - Towards a better understanding of variable and objective space through interactive exploration. Interactive Evolutionary Computation, Series of Workshops at the Genetic and Evolutionary Computing Conference (GECCO), http://www.ad-comtech.co.uk/Workshops.htm

Parmee I.C. (2002). Improving Problem Definition through Interactive Evolutionary Computation. Journal of Artificial Intelligence in Engineering Design, Analysis and Manufacture-Special Issue: Human-computer Interaction in Engineering Contexts 16(3).

Parmee I.C., Cvetkovic D., Watson A.H. Bonham C.R. (2000). Multi-Objective Satisfaction within an Interactive Evolutionary Design Enviroment. Journal of Evolutionary Computation 8 MIT Press, pp-197:222.

Parmee I.C. and Bonham C.R. (2000). Towards the Support of Innovative Conceptual Design Through Interactive Designer / Evolutionary Computing Strategies. Journal of Artificial Intelligence in Engineering Design, Analysis and Manufacture, **14**. Cambridge Press, pp 3-16.

Parmee I.C. (1996). The Maintenance of Search Diversity for Effective Design Space Decomposition using Cluster Oriented Genetic Algorithms (COGAs) and Multi-Agent Strategies (GAANT). Proceedings of 2nd International Conference on Adaptive Computing in Engineering Design and Control, Plymouth, UK; pp 128-138.

Webb E. (1997). MiniCAPS- A Simplified Version of CAPS for Use as a Research Tool. Unclassified Report BAe-WOA-RP-GEN-11313, British Aerospace.

Zitzler E., Deb K. and Thiele L. (2000), Comparison of Multi Objective Evolutionary Algorithms: Empirical Results, Evolutionary Computation, **8**(2), pp: 173-195.

Ziztler E., Laumanns M. and Thiele L, (2002). SPEA-II: Improving the Strength Pareto Evolutionary Algorithm for Multi Objective Optimisation. Evolutionary Methods for Design Optimisation and Control, CIMNE, Barcelona, Spain, pg 95-100.